Principal Factor Analysis of Stock Market Sentiment

DUAN, Xin

A Thesis Submitted in Partial Fulfillment

of The Requirements for the Degree of

Master of Philosophy

in

Economics

The Chinese University of Hong Kong

August 2007

Abstract

It has long been observed that investor sentiment has a crucial impact on the asset market. However, only a few studies are available due to the lack of a proper measure of sentiment. This thesis provides the first attempt to design a measure for investor sentiment. A composite index reflecting the stock market sentiment (SMS) is constructed by using the principal factor analysis. Factors such as the short sell volume, the Hong Kong Inter-bank Offered Rate (HIBOR), the RSI, the MFI, the turnover and the cross effects from related equity markets are considered.

A multivariate threshold model is estimated based on the daily return series of the Hang Seng Index (HSI) from 1998 to 2006. The composite index is used as the threshold variable to separate the unobserved states of the stock market. Our empirical results show that the stock market can be divided into three regimes: "optimistic market", "pessimistic market" and "neutral market". A trading strategy is proposed and encouraging results are found which support that strategy.

摘 要

長期觀察證實，投資者情緒對資產市場有著決定性的影響。但是，適當的情緒衡量方法的缺乏使可利用之研究少之又少。本論文嘗試設計一種投資者情緒衡量方法：使用主成分分析法構造一個反映股票市場壓力的綜合指數，其中主要因素包括市場賣空數量，香港銀行同業拆息，相對強弱指數，資金流向指數，交易量以及相關資本市場的交叉影響。

本文運用多元閾值模型估計了 1998 年至 2006 年恒生指數的每日回報。已構造的綜合指數作爲閾值變數來劃分股票市場。實證結果顯示，股票市場可被分爲樂觀市場，悲觀市場以及中立市場三個部分。在此基礎上，一項交易策略被提出且支援該策略的令人鼓舞的結果亦在本文中被找到。

Acknowledgement

I would like to express my deepest gratitude to my supervisor, Professor Chong Tai Leung, for his expert guidance and valuable suggestions in bringing this thesis to completion. I am also very grateful to Professor Du Julan and Professor Zhang Jun who spent their precious time in assessing my thesis and provided helpful comments and recommendations. Special thanks also go to many of my friends and colleagues for their support and encouragement.

TABLE OF CONTENTS

Chapter 1

Introduction

The nonlinear properties of economic and financial data have been widely studied in recent decades. A simple way to model nonlinear data is to use the threshold model of Tong (1980) or the regime switching model of Hamilton (1989). The two models classify each observation into one of the three regimes based on some threshold conditions[1]. These kinds of local-linear models have been performing very well in empirical studies[2]. The related estimation and inference theories have been extensively studied by Hansen (1996, 2000) and Hamilton (1989).

This thesis attempts to classify the stock market states via a newly developed method. The stock market is usually classified into bull or bear market. However, different classifications can be defined using different measures. Fabozzi and Francis (1977), Kim and Zumwalt (1979), and Chen (1982), define a bull market as a period when returns in any given month exceed a certain threshold value. Recently, Pagan and Sossounov (2000) define the bull and bear states based on cumulative price changes. Lunde and Timmermann (2004) classify the bull and bear states via a filter that tracks the movements between

[1] Tong (1980) uses one threshold variable to classify the data; Hamilton (1989) classifies the data according to some latent variables.

[2] See Henry et al. (2001), Potter (1995), Tsay (1998) and Dueker et al. (2003).

local peaks and troughs. Chen and Chong (2005) use a multiple threshold variables approach to define the bull and bear markets by combining price and turnover information.

Most of the time, it is not sufficient to determine the states of the stock market using a single variable. Misleading results might be obtained as some important information may possibly be neglected. To avoid this pitfall, Chong and Yan (2004), Chen and Chong (2005)[3] develop a threshold model with multiple threshold variables. A shortcoming of this kind of model is that the computation time increases exponentially with the number of threshold variables. Moreover, if the structures of the threshold variables are unknown, the asymptotic properties of the estimators will be complicated[4].

Fan and Yao (2003) suggests a parsimonious way to construct a composite index. They condense the information from all indicators into a single summary statistic. Such composite indices are not new to financial markets. Indeed, for many years, they have been used to identify market types and to forecast market performances[5]. Following Fan and Yao (2003), the present thesis develops a

[3] They try to design a way to find proper threshold variables from a set of candidate variables and determine the classification of regimes.

[4] see Chong and Yan (2004)

[5] For example, the El Paso index of coincident economic activity complied by the Federal Reserve Bank based on the nonagricultural employment, unemployment rate, inflation-adjusted wages and inflation-adjusted retail sales. Baker and Wurgler (2006) construct a composite index of stock market sentiment based on the common variation in six underlying proxies for sentiment: closed-end fund discount, NYSE share turnover, number and average first-day returns on IPOs,

stock market sentiment index to classify the states of the stock market.

This thesis uses the principal factor model to summarize information from various sources. The principal components estimator has been studied by Chamberlain and Rothschild (1983). It converges to the maximum likelihood estimator when the number of factors N increases. The inference theory related to the estimator has been studied by Bai (2003) and Bai and Ng (2006), who show that when the sample size is large, the space spanned by the latent factors can be consistently estimated from large sets of the observed information[6]. The principal factor analysis is an effective way to isolate the common factor even among a limited number of factors. It has been widely applied in asset pricing and co-movement studies in the field of finance[7].

In this thesis, a stock market sentiment index (SMS), which is assumed to be unobservable and determines the market status implicitly, is developed. Stock market commentators often refer to certain financial variables as market weather vanes. We examine a number of indicators, which are divided into five categories. The first part is the short sell volume, which is viewed as a bearish indicator. It has long been argued that short selling activities reflect downward expectation of

---

equity share in new issues and dividend premium. Blanco and Garber (1986), Weymark (1995, 1998) and Eichengreen et.al. (1995) form an exchange market pressure index (EMP) to measure the exchange market pressure.

[6] In this study, due to data limitation, a large sample of the informed variables cannot be obtained. As a result, the asymptotic results cannot be used.

[7] See Bai and Ng (2006)

the investors. When short sell volume becomes relatively large, the market is likely to fall. The second element is the HIBOR. This measure is frequently cited as a bearish indicator since higher interest rate drives money out of the stock market. In addition, two technical indicators are also included. Evidence of profitability has been found for some momentum-based trading strategies, such as strategies based on the Relative Strength Index (RSI) and the Money Flow Index (MFI). Thus, our pressure index incorporates those two indices. Moreover, due to the globalization trend and the high liquidity of the international financial markets, cross effects from the world's major markets, such as the US and the Japanese equity markets, are taken into consideration. Finally, liquidity is examined and treated as a bullish indicator. Based on the aforementioned factors, we employ the principal factor analysis to obtain the first principal component, which we define as the stock market sentiment index.

We apply the new sentiment measure to classify the Hong Kong stock market into different regimes. A multivariate threshold model is estimated for the return series of the Hang Seng Index (HSI) from 1998-2006 using the new sentiment measure as the threshold variable. Two shifts are identified and the market is classified into three regimes: an "optimistic regime" when the stock market sentiment index (SMS) exceeds the higher threshold value, a "pessimistic regime" when the SMS is smaller than the lower threshold value and a "neutral regime" when the SMS is in-between.

The thesis is organized as follows. Chapter 2 provides a literature review on some composite indices related to the market pressure. Chapter 3 discusses the stock market sentiment index, the construction method and the proxies used. In Chapter 4, the new index is applied to the Hong Kong stock market to test whether there are threshold effects on the return series of the market index. Chapter 5 concludes the study.

Chapter 2

Literature Review

2.1 Exchange Market Pressure Index

Exchange market pressure measures the total excess demand for a currency in international markets and it is defined as the exchange rate change that would have been required to remove this excess demand in the absence of exchange market intervention, given the expectations generated by the exchange rate policy actually implemented[8]. Generally, two methods are used to measure exchange market pressure.

The first method originates from Girton and Roper (1977). The asset market approach is used to measure the excess demand for a currency. The term "exchange market pressure" refers to the magnitude of the domestic money market disequilibrium that must be removed through changes in foreign exchange reserves or exchange rate. Girton and Roper (1977) derive their measure of exchange market pressure in a monetary model with two large interdependent economies and use it to quantify the volume of the central bank foreign exchange intervention to achieve any desired exchange rate target. Due to the assumption that the central bank does not change domestic credit to influence

[8] see Weymark (1998) p. 106

the exchange rate, the exchange market pressure (EMP) is a simple sum of percentage changes in the exchange rate and foreign exchange reserves of the domestic central bank.

Girton and Roper's (1977) original exchange market pressure formula has been used extensively[9]. The small open economy version of the Girton and Roper (1977)'s measure of exchange market pressure is formally derived in Connolly and Da Silveira (1979).

Roper and Turnovsky (1980) work along the line of Girton and Roper by introducing a stochastic small open economy IS-LM model to quantify the international excess demand for domestic currency. The model is based on two assumptions. The first is that excess demand is absorbed through a country's international reserves and the change in exchange rate. The second is that the intervention is allowed to take the form of changes in domestic credit as well as in reserves. The excess demand for money is therefore a linear combination of two unequally weighted components: the changes in the exchange rate and the money base. Following Girton and Roper (1977), Roper and Turnovsky (1980) define this excess demand as the exchange market pressure.

However, none of the aforementioned articles uses an independent

[9] see Burdekin and Burkett (1990), Hacche and Townend (1981), Kim (1985), Lee and Wohar (1991), Mah (1991,1995,1998), Modeste (1981), Pollard (1999), and Wohar and Lee (1992).

definition of the exchange market pressure to derive an expression of the exchange market pressure for economies with intermediate exchange rate systems. The first notable contribution to the independent definition is made by Weymark (1995), who constructs a model-dependent exchange market pressure index based on a general, model-independent definition of the exchange market pressure. His index is defined as the change in exchange rate which would be required to remove the excess demand for the currency in the absence of exchange market intervention, given the expectations generated by the exchange rate policy actually implemented[10]. Weymark (1998) illustrates the general applicability of the method and also extends the exchange market pressure index of Girton and Roper (1977) and Roper and Turnovsky (1980).

Spolander (1999) points out that a major shortcoming of the existing index is that it does not take into account the sterilization effect of the central bank's foreign exchange intervention. The liquidity effect of a foreign exchange intervention on the domestic money market's operation is partly or totally offset by central banks in the main industrial countries[11]. Therefore, it is important to take into account the realistic measurement of the exchange market pressure as well as the degree to which this pressure is removed by the central bank's foreign

---

[10] see Weymark (1995), p.278.

[11] For example, in the United States the Federal Reserve sterilizes the liquidity effect regularly and completely, so that a foreign exchange intervention would not lead to a change in the domestic money market and the domestic interest rate, compared to that which would have occurred in the absence of foreign exchange intervention. The case is generally the same in Finland.

exchange intervention. Spolander (1999) derives a measure of the pressure and the degree of intervention in the context of a small open-economy monetary model. Monetary policy is conducted through changes in the monetary base and the foreign exchange interventions. Different from other studies, the required parameters are estimated from the actual intervention data.

The second approach to measure the exchange market pressure is proposed by the critics of the model-dependent measure[12]. Eichengreen et al. (1995, 1996) argue that model-dependence is undesirable in an operational index given the tenuous connection between the exchange rate and economic fundamentals. As a result, a model-independent measure of EMP is derived. The new EMP is a weighted linear combination of the change in relevant interest rate differential, the percentage change in bilateral exchange rate and the percentage change in foreign reserves of the domestic central bank. The weights are chosen to equalize the conditional volatilities of EMP's constituent components. The index is model-independent as both the components of the index and the weights assigned to these components are not derived from a structural model of the economy[13].

Weymark (1995) defines EMP as

$$EMP_{it} = \Delta e_t + \eta \Delta r_t ,$$

[12] See model dependent measure developed by Girton and Roper's (1977), Roper and Turnovsky's (1980) and Weymark's (1995, 1997a, 1997b, 1998).

[13] See Weymark (1998) p. 118.

where

η is the elasticity $-\partial\Delta e_t / \partial\Delta r_t$

Δe_t is the change in the logarithm of the period t exchange rate expressed as the domestic currency cost of one unit of foreign currency.

$\Delta r_t = [h_t R_t - h_{t-1} R_{t-1}] / M_{t-1}$, where h_t is the money multiplier in period t, R_t is the stock of foreign exchange reserves in period t, and M_{t-1} is the inherited money stock in period t.

Eichengreen et al. (1995) also include the short-term interest rate differential in the measure. They define the EMP as follows,

$$EMP_{it} = \alpha_1 \%\Delta e_t + \alpha_2 \Delta(i_{it} - i_{US,t}) + \alpha_3 \%\Delta r_{it}$$

where

$\%\Delta e_t$ denotes the percentage change in the exchange rate of country i with respect to the U.S. dollar at time t.

$\Delta(i_{it} - i_{US,t})$ denotes the change in the difference between the short-term discount rate in country i and in the US at time t.

$\%\Delta r_{it}$ denotes the percentage change in the foreign exchange reserves of country i at time t.

α_is are the weights defined as the inverse of the standard deviations of the respective components.

2.2 The Sentiment Index

Empirical evidence has long suggested that investor sentiment can be used to explain the stock market. The topic has attracted a great deal of attention and studies with different measures of investor sentiment have been conducted over the last decades. In the following sub-section, we will shed light on these measures of investor sentiment and report the empirical results.

Many proxies for the sentiment are derived from financial markets. The most prominent one is proposed by Lee et al. (1991), who argue that closed-end fund discounts are a measure of the individual investor sentiment. Since both closed-end fund and smaller stocks are mainly owned by individual investors, the sentiment affects the returns of smaller stocks in the same way as it influences the closed-end funds. The results of Lee et al (1991) are partially supported by Swaminathan (1996)[14] and Neal and Wheatley (1998)[15].

Using data from a major discount brokerage house, Kumar and Lee (2002) define a measure of individual investor sentiment as a systematic component which is contained in the buy-sell imbalance in individual investors' trades and is uncorrelated with overall market movements. They find that the sentiment has incremental explanatory power for small stocks, value stocks, stocks with low institutional ownership and stocks with lower prices. On the other hand, the explanatory power appears to be affected by factors associated with arbitrage costs. In addition, Kumar and Lee (2004) find that trading activities of retail investors can explain the co-movement in stock returns.

Another market sentiment measure is constructed by Baker and Stein (2003). They form a composite stock market investor sentiment based on the common variation in six underlying proxies: the closed-end fund discount, the NYSE share turnover, the number of IPOs, the average first-day returns on IPOs, the equity share in new issues and the dividend premium. A long-term investor sentiment index is constructed using the historical yearly data from 1934 to 2004.

[14] Swaminathan also find that closed-end fund discounts forest future excess returns on small firms. However, he also finds that discount could forecast future earnings growth rate on small firms, thus it "may be a proxy of individual investors' rational expectations about future economic conditions and/or their risk aversion to macroeconomic risks" instead of investor sentiment.

[15] Neal and Wheatley find that fund discount predict equity return.

They also find that when sentiment is high, stocks attractive to optimists and speculators are unattractive to arbitrageur while, conditional on low sentiment, these cross-sectional patterns completely reverse.

Recently, Brown and Cliff (2004), Wang et al. (2006), Dorn et al. (2004) and Schmitz et al. (2006) provide additional sentiment measures that are methodologically related[16]. The first two studies test various proxies for investor sentiment within a vector autoregressive (VAR) model to investigate the mutual influence between those measures and market variables such as returns and volatility. Brown and Cliff (2004) show that the direct measures of investor sentiment based on survey data are related to the commonly used indirect investor sentiments. Thus by employing the Kalman filter and the principal component analysis, they are able to isolate the common factors from the indicators for monthly and weekly data and to separate the measures for institutional and individual investor sentiment in the weekly data. For the composite measures, strong evidence of co-movement with the market is found. However, contrary to some aforementioned studies, short-run predictability of them in returns is hardly supported using a VAR model. And a trading strategy accordingly does not seem to be profitable. Supporting and expanding Brown and Cliff's study, Wang et al (2006) also construct two different sentiment indicators using the daily and weekly data separately, the daily sentiment is

[16] See Wang, Keswani and Taylor; Dorn, Huberman and Sengmueller; Schmitz, Glaser and Weber.

constructed from the S&P 100 put-call trading volume ratio (PCV), the S&P put-call open interest ratio (PCO) and the NYSE ARMS[17] index while the weekly sentiment is extracted from the PCO and the PCV. Two additional survey based sentiments[18] are used as well. In their study, the sentiment measures are shown to be caused by realized volatility and returns rather than the other way around. In addition, returns may be useful in predicting realized volatility. Dorn et al (2004) detect a tendency of herding within a group of individual investors from an online broker. They use the buy ratio as a sentiment variable and estimate its forecasting power for stock returns with a VAR model. It is found that there exists a positive relation between net trading by individuals and returns. The results are especially strong in the very short-run[19].

Schmitz et al. (2006) are the first to analyze investor behavior in the warrant market. They construct a daily measure of individual investor sentiment using a data set from a big German online brokerage company between Jan. 1997 to April 2001. The mutual relationship between the sentiment measure and daily stock returns is tested in a VAR model and with Granger-causality test. A short-term[20] mutual influence is found: sentiment influences stock returns positively while returns have a negative influence on sentiment.

---

[17] The ARMS index is named after its creator Richard Arms, it is the ratio of the average volume of advancing versus declining issues.

[18] Survey based sentiment will be discussed later.

[19] Very short run is the daily frequency.

[20] One or two trading days.

The sentiment measures described above are based on financial markets. Another category of sentiment measure is derived from survey-based proxies among institutional investors, individual investors, managers or consumers. For example, Otoo (1999) studies the survey-based Michigan consumer confidence index and finds that there is a strong positive relationship between the consumer sentiment and stock prices. Clarke and Statman (1998) examine a sentiment indicator published by Investors Intelligence (II) which is based on the number of newsletters that are bullish, neutral, or bearish. It is found that in the time periods from 1963 to 1985 and from 1963 to 1995, sentiment does not influence stock market returns over 4 weeks, 26 weeks, and 52 weeks measures while it is influenced by returns lagged over these horizons. Wang et al. (2006) also use two sentiment indicators on a weekly basis that are compiled from surveys by the AAII[21] and II.

[21] American Association for Individual Investors

Chapter 3

Stock market sentiment

3.1 Data

The stock market sentiment is designed to measure the market sentiment. The first step is to gather proxies for the sentiment measure. What proxy variables should be included is always a controversial issue. In this paper, we consider a number of proxies and form the composite index based on their first principal component. The factors included are the short-sell ratio, the turnover, the HIBOR, technical indicators including the RSI and the MFI, and the cross effect from USA and Japanese equity markets. A brief description of each one follows.

Turnover

There is an old Wall Street adage saying that: "It takes volume to make prices move." Although one can question the asserted causality, numerous empirical findings support this so-called positive relationship between trading volume and stock prices. For example, Ying (1966) finds that a small trading volume is typically followed by a fall in price while a large trading volume is typically associated with a rise in price. In general, volume is higher in bull markets and lower in bear markets.

The daily turnover data is obtained from DataStream. The ratio of 250-day average line and 20-day average line is calculated to smooth the data, thus the number of effective observations is about 1800. The turnover ratio is defined as follows and a time series graph with average lines is shown in Figure 2.

* * * * * * * * * * * * * * * FIGURE 2 HERE * * * * * * * * * * * * * *

$$Rav_t = 100 \times \frac{VMA10_t}{VMA250_t}$$

where $VMA250_t$ are the average price and turnover for the past 250 trading days, while $VMA10_t$ are average price and turnover for the past 10 days[22].

Short Sellings in the Hong Kong Stock Market

Figlewski (1981) studies the relationship between the level of short interest and subsequent stock returns. He argues that abnormal returns are often due to temporary sentiment brought about by short-selling activities around the event dates. In practice, short selling usually reflects downward expectation on the part of the investors. In the Hong Kong stock market, short sellings are allowed only for a small number of stocks. During recent years, the highly developed derivative market has provided a convenient platform for short sellings. We

[22] The 10-day average is used here for smoothing, while he 250-day average line is usually used to differentiate between bull and bear markets.

define the daily short-sell turnover ratio as the number of shares being sold short to the total number of shares traded on that day[23]. Those data are available on a daily basis from the CEIC database.

$$SR_t = \frac{short \quad sell \quad vol_t}{turnover_t}$$

From Figure 3, the ratio seems to be more stationary after adjusting for the turnover.

* * * * * * * * * * * * * * * FIGURE 3 HERE * * * * * * * * * * * * * *

Hong Kong Inter-bank Offered Rate (HIBOR)

HIBOR is the inter-bank interest rate, which reflects the capital cost of investments. The terms of the deposits vary from overnight to one year. In practice, a high HIBOR is viewed as a bearish indicator and investors leave the stock market when HIBOR is high. Moreover, a higher HIBOR increases the cost of investment and hence lower the profits of the listed companies, which will be reflected in their share prices.

The daily HIBOR obtained from the CEIC database is plotted in Figure 4.

[23] The aim of the division is to move the trend that short sell volume increases with market turnover.

The HIBOR appears to vary in the 0-10% range.

* * * * * * * * * * * * * * * FIGURE 4 HERE * * * * * * * * * * * * * * *

Technical Indicators

Money flow reflects the dollar value of a day's trading. The money flow index (MFI) is an oscillator calculated over an N-day period, ranging from 0 to 100, showing the money flow on up days as a percentage of the total of up and down days. The calculation is as follows. The daily typical price is the average of high, low and close,

$$Typical \quad Price = \frac{high + low + close}{3}$$

Money flow is the product of typical price and the volume on that day.

$$Money \quad Flow = Typical \quad Price \times Turnover$$

The total of the money flow over the given N days is computed. In this study, N is set to be 30[24]. If typical price today is higher than that of yesterday, it is considered positive money. For a 30-day average, the sum of all positive money for those 30 days is the positive money flow. We define

[24] A swing trader might prefer 14-day periods, while an investor may prefer 30-day periods.

$$Money \quad Ratio = \frac{Positive \quad Money \quad Flow}{Negative \quad Money \quad Flow}$$

The money flow index (MFI) can be calculated as follows:

$$MFI = 100 - \frac{100}{1 + money \quad ratio}$$

The MFI can be expressed equivalently in another form as:

$$MFI = 100 \times \frac{Positive \quad money \quad flow}{Positive \quad money \quad flow + Negative \quad money \quad flow}$$

The MFI is used as an oscillator. A value of 80 is generally considered to characterize an overbought market and a value of 20 is generally associated to an oversold market[7]. Divergences between MFI and price action are also considered significant. For instance, if the price of a stock makes a new rally high but the MFI high is less than its previous high, it may indicate that a weak advance, and the market trend is likely to reverse. The MFI is constructed in a similar fashion to the relative strength index.

Up days are the days when the market is dominated by buyers, while down days represents the days when sellers dominate. An excessive proportion in one

[7] See, for example, http://www.chartfilter.com/reports/c25.htm.

direction or the other is interpreted as an extreme and a price reversal is likely to occur.

The Relative Strength Index (RSI) reflects whether the market is oversold or overbought. Usually, an RSI of 80 implies that the market is overbought, while an RSI of 20 implies the market is oversold. The 14-day RSI is defined as follows:

$$RSI(14)_t = 100 \times \frac{\sum_{i=1}^{14} (P_{t-i} - P_{t-i-1})_+}{\sum_{i=1}^{14} \left| P_{t-i} - P_{t-i-1} \right|}$$

Figure 5 plots the RSI and MFI.

* * * * * * * * * * * * * * * FIGURE 5 HERE * * * * * * * * * * * * * * *

Cross Effect from USA and Japanese Equity Markets

The highly integrated global equity market has brought obvious cross-listing effects. The market performances in USA and Japan can usually be used to forecast the performances of the Hong Kong stock market.

Since the New York Exchange trades in a different time zone than the Hong Kong Exchange, the lag return of the market needs to be taken into account. The

S&P 500 is used to calculate the return of the US market, while the NIKKEI 225 is used for the Japanese market. The two proxies are defined as follows:

$$SP_t = (S\&P_{t-1} - S\&P_{t-2}) / S\&P_{t-2}$$

$$JAP_t = (NIK_t - NIK_{t-1}) / NIK_{t-1}$$

Figure 6 shows the daily return series of NIKKEI 225 and S&P 500 obtained from DataStream.

* * * * * * * * * * * * * * * FIGURE 6 HERE * * * * * * * * * * * * * *

The descriptive statistics of the adjusted data are summarized in Table 1.

* * * * * * * * * * * * * * * TABLE 1 HERE * * * * * * * * * * * * * *

Market Performance

About 2,000 daily prices of the Hang Seng index from Jan.1998 to Aug. 2006 are obtained from DataStream. The return series is defined as the log difference of the index. The time series plot is given in Figure 1.

* * * * * * * * * * * * * * * FIGURE 1 HERE * * * * * * * * * * * * * *

3.2 Methodology

The following multivariate model is considered with p_t as the log price of the asset at time t, and z_t as the threshold variable which determines the latent states of the market:

$$R_t = \begin{cases} f_1(R_{t-1}, R_{t-2}, \ldots; \varepsilon_{1t} | \theta_1) \ \ if & z_t \le r_1 \\ f_2(R_{t-1}, R_{t-2}, \ldots; \varepsilon_{2t} | \theta_2) \ \ if & r_1 \le z_t \le r_2 \\ \ldots\ldots & \\ f_n(R_{t-1}, R_{t-2}, \ldots; \varepsilon_{nt} | \theta_n) \ \ if & r_n \le z_t \end{cases}$$

$f_i(.)s$ are well-defined functions such that $f_i(.) \ne f_j(.)$ for $i \ne j$, θ_i are finite-dimensional parameters, ε_{it} denote the noise terms and $r_1 \ldots r_n$ are real numbers. In time series, we usually set the function $f_i(.)$ as a linear function in each regime. Thus, the model incorporates the multivariate threshold autoregressive model.

Tsay (1998) constructs a test statistic for detecting threshold nonlinearity and proposes a procedure to estimate the model by assuming that the threshold variable z_t is given. We extend the model by assuming that the threshold variable z_t is unobservable and incorporates lots of variables x_{it} $(i = 1, 2, \ldots, p)$, which are available in our data set.

Let

$$z_t = \lambda_i x_{it} + e_{it}$$

be the factor representation of the data. We assume that z_t is a scale number, i.e., each proxy x_{it} just includes one common component as well as idiosyncratic components. This assumption allows us to focus on a single threshold variable model.

To extract the common component from x_{it}, the principal factor approach, which has been widely adopted to obtain the common factor from a lot of variables, is used[26].

We first provide a brief introduction to the principal factors analysis.

Definition $X=(x_1,x_2,...,x_p)=\begin{pmatrix} x_{11}, & \cdots & ,x_{1p} \\ \vdots & \cdots & \vdots \\ x_{T1}, & \cdots & ,x_{Tp} \end{pmatrix}$ *is* $T\times p$ *matrix.* $\sum$ *is the covariance matrix of* X. $Z_i = a_{i1}x_1 + a_{i2}x_2 + ... + a_{ip}x_p$ $(i=1,...,T)$ *is the* $i-th$ *principal factor of* X *if*

(i) $a_\gamma^{'} a_\gamma = 1$ $(\gamma = 1,...,p)$;

(ii) For any $i>1, a_i \sum a_j = 0$ $(j=1,...,i-1)$;

(iii) $Var(Z_i) = \max(a'X)$ *for all* a *satisfying (i) and (ii).*

[26] see Bai and Ng (2006) and Baker and Wulger (2006).

Using the above definition, we have the following theorem:

Theorem *Let* $X=(x_1,x_2,...,x_p)$ *be a* $T\times p$ *matrix.* $\sum$ *is the covariance matrix of* X, *and* $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ *are the eigen values of* $\sum$ *.*a_i *is the corresponding eigen vector for* λ_i. *Then* $z_i = a_i' X$ *is the ith principal factor of* $X(i=1,2,...,p)$.

From the above theorem, to calculate z_t, we only need to find the eigen vector associated with the largest eigen value of the matrix $\sum = XX/Tp$, where $X=(x_1,x_2,...,x_p)$ is a $T\times p$ matrix. The proof is given in the Appendix.

3.3 Estimated Results

We form the composite index to capture the common component of the following factors: turnover, past price momentum, money flow, short term interest, short sell volume, and the cross market performances of US and Japan. The weight of each proxy is calculated by using the principal factor analysis.

$$SMS_t = -5.22 + 1.34Rav_t + 0.03RSI_t + 0.053MFI_t - 1.04HIBOR_t - 16.65SR_t + 2.69SP_t + 2.25JAP_t$$

where $Rav_t, RSI_t, MFI_t, HIBOR_t, SR_t, SP_t, JAP_t$ represent the information from turnover, past price momentum, money flow, short term interest, short sell volume and cross market performances of US and Japan respectively.

To gain some intuition on the stock market sentiment index, Figure 7 shows the time series graph of the estimated stock market sentiment index.

* * * * * * * * * * * * * * FIGURE 7 HERE * * * * * * * * * * * * * * *

According to the results, the turnover is, as expected, a bullish indicator which is positively related with price movement: it is higher in bull market while lower in bear market.

Both the RSI and MFI positively affect the stock market sentiment index. The RSI reflects the past price momentum in recent trading days. A larger RSI indicates that investors have a higher incentive to buy stocks. The MFI is the money flow index. When the money flow is positive, the market is expected to have a positive price movement.

The HIBOR has a negative effect on the stock market sentiment index. This bearish indicator is the short-term interest rate, which reflects the opportunity cost of investment on the stock market. A higher HIBOR will drive money out of the stock market.

Our results confirm that short-sell reflects the short-term downward pressure of the market. When there is large volume of short sell, market is likely to fall.

The performances of the US and Japanese equity markets have positive effect on the Hong Kong stock market due to the globalization of the international equity market.

Chapter 4

Application to the Hong Kong stock market

4.1 Threshold Model Estimation

It has long been argued that the behavior of the stock prices is consistent with a nonlinear data-generating process. A growing body of studies on the single threshold variable model has been proposed to describe the non-linearity found in the stock price series. For example, Tong (1983) uses the STAR model to fit the IBM stock price. Li and Lam (1995) investigate the asymmetric behavior of stock prices in bear and bull markets by using a threshold type ARCH model. Recently, Shively (2003) employs a three-regime threshold random-walk model to fit the daily data of the CAC 40, the DAX 30 and other indices[27].

In this study, a multivariate TAR model is estimated using the stock market sentiment index as the threshold variable. The model can be written as follows:

$$R_t = \begin{cases} a_0 + a_1 R_{t-1} + a_2 R_{t-2} + \ldots + a_{p1} R_{t-p1} + \varepsilon_{1t} & if \quad SMS_t \le r_1 \\ \beta_0 + \beta_1 R_{t-1} + \beta_2 R_{t-2} + \ldots + \beta_{p2} R_{t-p2} + \varepsilon_{2t} & if \quad r_1 \le SMS_t \le r_2 \\ \phi_0 + \phi_1 R_{t-1} + \phi_2 R_{t-2} + \ldots + \phi_{p3} R_{t-p3} + \varepsilon_{3t} & if \quad r_2 \le SMS_t \end{cases}$$

[27] CAC 40 is a French stock market index while DAX 30 is a Blue Chip stock market index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange.

According to the results, the SMS varies between –4 and 4, a threshold value of –0.78 is firstly found between (-4, 4), another value of 1.42 is then found between (-0.78, 4).

To test whether there are threshold effects, the likelihood ratio test applied in Chong and Yan (2004), Chen and Chong (2005) is employed here. To obtain the critical values of the threshold variables, bootstrap method is used. The test results are as follows:

Table 2: The Threshold Test Result

Threshold	-0.78	1.42
Statistics Value	13.07	25.48
Bootstrap Critical Value	9.62	11.02

Both threshold values are significant from the test results[28]. Our estimation results are:

$$R_t = \begin{cases} -0.0006 - 0.047R_{t-1} + \varepsilon_{1t} & if \quad SMS_t \le -0.78 \\ 0.0007 + 0.048R_{t-1} + \varepsilon_{2t} & if \quad -0.78 \le SMS_t \le 1.42 \\ 0.0015 + 0.17R_{t-1} - 0.21R_{t-2} + \varepsilon_{3t} & if \quad 1.42 \le SMS_t \end{cases}$$

[28] The observed test values of both threshold are larger than their bootstrap critical values at the 5% level.

Based on the estimation results, the stock market is classified into three regimes: the "optimistic market" when SMS_t is larger than 1.42; the "neutral market" when SMS_t is between -0.78 and 1.42; and the "pessimistic market" when SMS_t is smaller than -0.78. The optimistic market is likely to have a positive momentum in the near future while the situation is opposite for the pessimistic market. The neutral market is neutral to downward or upward movement.

To justify the classification, the SMS and Hang Seng Index are compared. In Figure 8, the SMS can usually be found on a peak point when the Hang Seng Index is at a trough, followed by a significant increase afterwards and vice versa.

* * * * * * * * * * * * * * * FIGURE 8 HERE * * * * * * * * * * * * * * *

4.2 Trading Strategy

In this part, a trading strategy is designed according to the regimes of the market. We assume that investors buy stocks in optimistic market, sell stocks in pessimistic market and hold cash in the neutral market to earn HIBOR returns.

Buy if $SMS_t > 1.42$,

Hold if $SMS_t \in [-0.78, 1.42]$,

Sell if SMS_t <-0.78.

To see whether the strategy can produce significant results, the trading-rule returns are calculated. We find that an average return of 0.00091 is expected in the optimistic market and -0.00050 in pessimistic market, which conforms to our classification of the market. In a neutral market, an average HIBOR return of 2.95% per year is earned. The average daily return obtained in stock market is 0.00064. For example, if one holds a contract of the Hang Seng Index futures, if the current HSI is 15,000, about 9.6 points can be earned each day, and since the payoff for each point in a futures contract is 50 HK dollar, 480 HK dollar can be earned each day per contract.

Chapter 5

Conclusion

The classification of the stock market states is an important issue. To fully reflect the information of the market, a composite index is constructed. The index is a linear combination of six factors[29] whose weights are obtained via the principal factor analysis. The new composite index is employed to describe the stock market sentiment (SMS), which reflects the investor sentiment. Our empirical results are appealing as each individual factor is found to have the expected sign: the turnover, the values of technical indicators and the performances of other equity markets are positively related with the SMS, while the short selling volume and the HIBOR have negative relationship with the SMS.

We then apply this new composite index to the Hong Kong stock market. A threshold model is estimated to identify the unobserved states of the Hong Kong stock market using the pressure index as the threshold variable. The results show that there are two thresholds in the Hang Seng Index from 1998 to 2006, which divide the market into three regimes. The optimistic market is also shown to have a significantly higher return than the pessimistic market. Based on the observation, a trading strategy is designed and a positive return is earned.

---

[29] Short sell volume, HIBOR, technical indicators, turnover volume and the cross effect from US and Japanese equity market.

An extension of this thesis would be to include more information into the construction of the stock market sentiment index, such as the premium of the closed-end fund or the future market premium. Based on the theoretical result from Bai and Ng (2006), the estimation of the common factor will be consistent if the number of proxies is large. However, it is beyond the scope of this paper and is left for further research.

Appendix: Principal Component

(1) $X=(x_1,x_2,...,x_p)$, $\sum$ is the $covariance$ $matrix$ of X

To find a_1 such that maximize $\mathrm{var}(a_1^{'}X)$ over all linear combinations of X subject to $a_1^{'}a_1=1$ $\Leftrightarrow$ maximize $a_1^{'}\Sigma a_1$ subject to $a_1^{'}a_1=1$.

Using the lagrangian method

L= $a_1^{'}\Sigma a_1-\lambda_1(a_1^{'}a_1-1)$;

a_1 satisfies $2(\Sigma-\lambda_1 I)a_1=0$

$\Rightarrow a_1^{'}\Sigma a_1=\lambda_1$, i.e. $\Sigma a_1=\lambda_1 a_1$;

so $|\Sigma-\lambda_1 I|=0$;

λ_1 is the largest eigen value of Σ;

a_1 is the eigen vector corresponding to λ_1;

and the first principal component $Z_1=a_1^{'}X$;

(2) Find a_2 that maximize $\mathrm{var}(a_2^{'}X)$ subject to $a_2^{'}a_2=1$ and $a_2^{'}\lambda_1 a_1=a_2^{'}\Sigma\, a_1=0$ (i.e. $a_1^{'}X$, $a_2^{'}X$ are uncorrelated)

Using the lagrangian method

$L=a_2^{'}\Sigma a_2-\lambda_2(a_2^{'}a_2-1)-2va_2^{'}\Sigma a_1$

$\Rightarrow 2\Sigma a_2-2\lambda_2 a_2-2v\Sigma a_1=0$

$\Rightarrow a_1^{'}\Sigma a_2-2\lambda_2 a_1^{'}a_2-2va_1^{'}\Sigma a_1=0$

$(\because a_2^{'}\Sigma a_1=a_2^{'}\lambda_2 a_1 \quad \therefore a_2^{'}a_1=0\ )$

$\Rightarrow v=0$;

$a_2^{'}$ is the eigenvector corresponding to λ_2 which is the second largest value var($a_2{}'X$)= λ_2 .

(3) By analogy, we can obtain the *k*-th $Z_T = a_T{}'X$

a_T is orthogonal to $a_1, a_2, \ldots, a_{T-1}$.

References

1. Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica* 71(1), 135-171.

2. Bai, J. and S. Ng (2006). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica* 74 (4), 1133-150.

3. Baker, M. and J. C. Stein (2002). Market Liquidity as A Sentiment Indicator, *NBER Working Papers* 8816.

4. Baker, M. and J. Wurgler (2006). Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance* 61 (4), 1645-1680.

5. Barberis, N, A. Shleifer and R. Vishny (1998). A Model of Investor Sentiment. *Journal of Financial Economics* 49, 307-343.

6. Bodurtha, J.N., D.S. Kim and C.M.C. Lee (1995). Closed-End Country Funds and U.S. Market Sentiment. *The Review of Financial Studies* 8 (3), 879-918.

7. Brown, G. W. and M. T. Cliff (2004). Investor Sentiment and the Near-Term Stock Market. *Journal of Empirical Finance* 11, 1-27.

8. Brown, G. W. and M. T. Cliff (2005). Investor Sentiment and Asset Valuation. *Journal of Business* 78, 405-440.

9. Chamberlain, G. and M. Rothschild (1983). Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets. *Econometrica*, 51,5,1281-1304.

10. Chang, E. C., J. W. Cheng and Y. H. Yu (2004). Short-Sales Constraints and Price Discovery: Evidence from the Hong Kong Market. *Working paper*.

11. Chen, H. Q. and T. L. Chong (2005). Threshold Autoregressive Model with Multiple Threshold Variables. *Working paper, the Chinese University of Hong Kong*.

12. Chen, H. Q. and T. L. Chong (2006). Asymptotic Theory for Multi-Threshold-Variable Autoregressive Models. *Working paper, the Chinese University of Hong Kong*.

13. Chen, H. Q., T. L. Chong and M. Hinich (2006). A New Approach to Estimate the Regime Switching Model of Stock Prices. *Working paper, the Chinese University of Hong Kong*.

14. Chen, R. and R.S. Tsay (1993). Functional-Coefficient Autoregressive Models. *Journal of the American Statistical Association*, 88, 298-308.

15. Chong, T. L. and K. M. Yan (2004). Threshold Model with Multiple Threshold Variables and Application to Financial Crises. *Working paper*, *The Chinese University of Hong Kong*.

16. Clarke, R and M. Statman (1998). Bullish or Bearish?. *Financial Analysts Journal,* 54 (2), 63–72.

17. Connolly, M. and J. Dantas da Silveira (1979). Exchange Market Pressure in Postwar Brazil: An Application of the Girton-Roper Monetary Model. *American Economic Review* 69 (3), 448-454

18. Cornelli, F., D. Goldreich and A. Ljungqvist (2006). Investor Sentiment and

Pre-IPO Markets. *The Journal of Finance* LXI (3), 1187-1216

19. Dueker, M., S. Martin and F. Spagnolo (2003). Contemporaneous Threshold Autoregressive Models: Estimation, Forecasting and Rational Expectations Application. *Working paper, Federal Reserve Bank of St. Louis*.

20. Dorn, D., G. Huberman and P. Sengmueller (2004). Do Individual Investors Move together and thereby Affect Prices, *Working Paper, University of Drexel*.

21. Eichengreen, B., A. K. Rose and C. Wyplosz (1995). Exchange Market Mayhem. *Economic Policy* 21, 249-312

22. Eichengreen, B., A. K. Rose and C. Wyplosz (1996). Speculative Attacks on Pegged Exchange Rates: An Empirical Exploration with Special Reference to the European Monetary System. In The New Transatlantic Economy, eds: M.B. Canzoneri and W.J. Ethier and V. Grille, Cambridge University Press, Cambridge, 191-235

23. Elton, E. J., M. J. Gruber and J. A. Busse (1998). Do Investor Care About Sentiment?. *The Journal of Business* 71(4), 477-500.

24. Fabozzi, F. J. and J. C. Francis (1977). Stability Tests for Alphas and Betas over Bull and Bear Market Conditions. *Journal of Finance* XXX 11(4), 1093-1099.

25. Fan, J. Q. and Q. Yao (2003). Nonlinear Time Series, Springer, 313-393.

26. Girton, L. and D. Roper (1977). A Monetary Model of Exchange Market Pressure Applied to the Postwar Canadian Experience. *American Economic*

Review 67, 537-548.

27. Haggan, V. and T. Ozaki (1981). Modeling Nonlinear Vibrations Using an Amplitude-dependent Auto-regressive Time Series Model. *Biometrika* 68, 189-196

28. Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57(2), 357-384.

29. Hansen, B. E. (1996). Inference when a Nuisance Parameter is Not Identified Under the Null Hypothesis. *Econometrica* 64, 413-430.

30. Hansen, B. E. (1999). Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference. *Journal of Econometrics* 93, 345-368.

31. Hansen, B. E. (2000). Sample Splitting and Threshold Estimation. *Econometrica* 68 (3), 575-603.

32. J Bradford, D. L., A. Shleifer, L. H. Summers and R. J. Waldmann (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy* 4, 703-738

33. James, C. and R. O. Edmister (1983). The Relation between Common Stock Returns Trading Activity and Market Value. *Journal of Finance* 38, 1075-1086

34. Jansen, W. J. and N. J. Nahuis (2003). The Stock Market and Consumer Confidence: European Evidence. Economics Letters 79(1), 89-98.

35. Kaizoji, T., S. Bornholdt and Y. Fujiwara (2002). Dynamics of Price and

Trading Volume in A Spin Model of Stock Markets with Heterogeneous Agents. *Physica* A 316, 441-452.

36. Kaniel, R., G. Saar and S. Titman (2004). Individual Investor Sentiment and Stock Returns, *Working Paper*

37. Karpoff, J. M. (1987). The Relation Between Price Changes and Trading Volume: A Survey, *The Journal of Financial and Quantitative Analysis* 22 (1), 109-126.

38. Kim, M. K. and J. K. Zumwalt (1979). An Analysis of Risk in Bull and Bear Markets, *Journal of Financial and Quantitative Analysis* XIV (5), 1015-1024.

39. Kumar, A., and C. M. C. Lee (2002). Individual Investor Sentiment and Comovement in Small Stock Returns. *Working paper, Cornell University.*

40. Kumar, A. and C. M. C. Lee (2004). Retail Investor Sentiment and Return Comovements. *The Journal of Finance* 61(5), 2451-2486.

41. Kyle, A. S. (1985). Continuous Auctions and Insider Trading, *Econometrica* 53(6), 1315-1336.

42. Lakonishok, J., A. Shleifer and R. W. Vishny (1992). The Impact of Institutional Trading on Stock Prices. *Journal of Financial Economics* 32, 23-43.

43. Lee, C. M. C., A. Shleifer and R. H. Thaler (1991). Investor Sentiment and the Closed-End Fund Puzzle. *The Journal of Finance* 46(1), 75-109.

44. Lee, C. M. C. and B. Swaminathan (2000). Price Momentum and Trading

Volume. *The Journal of Finance* 55, 2017-2069.

45. Li, W. K. and K. Lam (1995). Modeling Asymmetry in Stock Returns by a Threshold Autoregressive Conditional Heteroscedastic Model. *Statistician* 44, 333-341.

46. Liu, S. M. (2006). Investor Sentiment and Stock Market Liquidity. *Working paper, University of Texas*.

47. Lunde, A. and A. G. Timmermann (2004). Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets. *Journal of Business and Economic Statistics* 22(3), 253-279.

48. Macedo, Jorge Braga de, José. Braz, Luís Brites. Pereira and Luís C. Nunes (2004). Exchange Market Pressure and the Credibility of Macau's Currency Board, Universidade Nova de Lisboa, *Working paper*.

49. Morck, R., A. Shleifer, R. W. Vishny, M. Shapiro and J. M. Poterba (1990). The Stock Market and Investment: Is the Market a Sideshow?. *Brookings Papers on Economic Activity* 1990(2), 157-215

50. Neal, R and S. M. Wheatley (1998). Do Measures of Investor Sentiment Predict Returns?. *The Journal of Financial and Quantitative Analysis* 33(4), 523-547.

51. Otoo, M.W. (1999). Consumer Sentiment and the Stock Market. *FEDS Working Paper* 99-60.

52. Pagan, A. R. and K. A. Sossounov (2003). A Simple Framework for Analyzing Bull and Bear Markets. *Journal of Applied Econometrics* 18,

23-46.

53. Potter, S. E. (1995). A Nonlinear Approach to U.S. GNP. *Journal of Applied Econometrics* 10, 109-125.

54. Qiu, L. and I. Welch (2004). Investor Sentiment Measures, *NBER Working Papers* 10794.

55. Roper, D. E. and S. J. Turnovsky (1980). Optimal Exchange Market Intervention in a Simple Stochastic Macro Model. *Canadian Journal of Economics, Canadian Economics Association* 13(2), 296-309.

56. Schmitz, P., G. Markus and M. Weber (2006). Individual Investor Sentiment and Stock Returns - What Do We Learn from Warrant Traders?. *Sonderforschungsbereich 504, Universität Mannheim* (06-12).

57. Shively, A. (2003), The Nonlinear Dynamics of Stock Prices. *Quarterly Review of Economics and Finance* 43, 505-517.

58. Spolander, M. (1999). Measuring Exchange Market Pressure and Central Bank Intervention. *Bank of Finland Studies, working paper* E:17

59. Swaminathan, B. (1996). Time Varying Expected Small Firm Returns and Closed-End Fund Discounts. *Review of Financial Studies* 9(3), 845-887.

60. Tong, H. and K. S. Lim (1980). Threshold Autoregression, Limit Cycles and Cyclical Data, *Journal of the Royal Statistical Society* Ser.B, 42(3), 245-292.

61. Tsay, R. S. (1998). Testing and Modeling Multivariate Threshold Models, *Journal of the American Statistical Association* 93, 231-240.

62. Wang, Y. H., A. Keswani and S. J. Taylor (2006). The Relationships Between

Sentiment, Returns and Volatility. *International Journal of Forecasting* 22(1), 109-123.

63. Weymark, D. N. (1995). Estimating Exchange Market Pressure and the Degree of Exchange Market Intervention for Canada. *Journal of International Economics* 39, 273-295.

64. Weymark, D. N. (1997a). Measuring the Degree of Exchange Market Intervention in a Small Open Economy. *Journal of International Money and Finance* 16(1), 55-79.

65. Weymark, D. N. (1997b). Measuring the Degree of Exchange Market Intervention in Interdependent Economies: A Two-Country Model. *Review of International Economics* 5(1), 72-82.

66. Weymark, D. N. (1998). A General Approach to Measuring Exchange Market Pressure. *Oxford Economic Papers* 50(1), 106-121

67. Ying, C. C. (1966). Stock Market Prices and Volumes of Sales. *Econometrica* 34(3), 676-685.

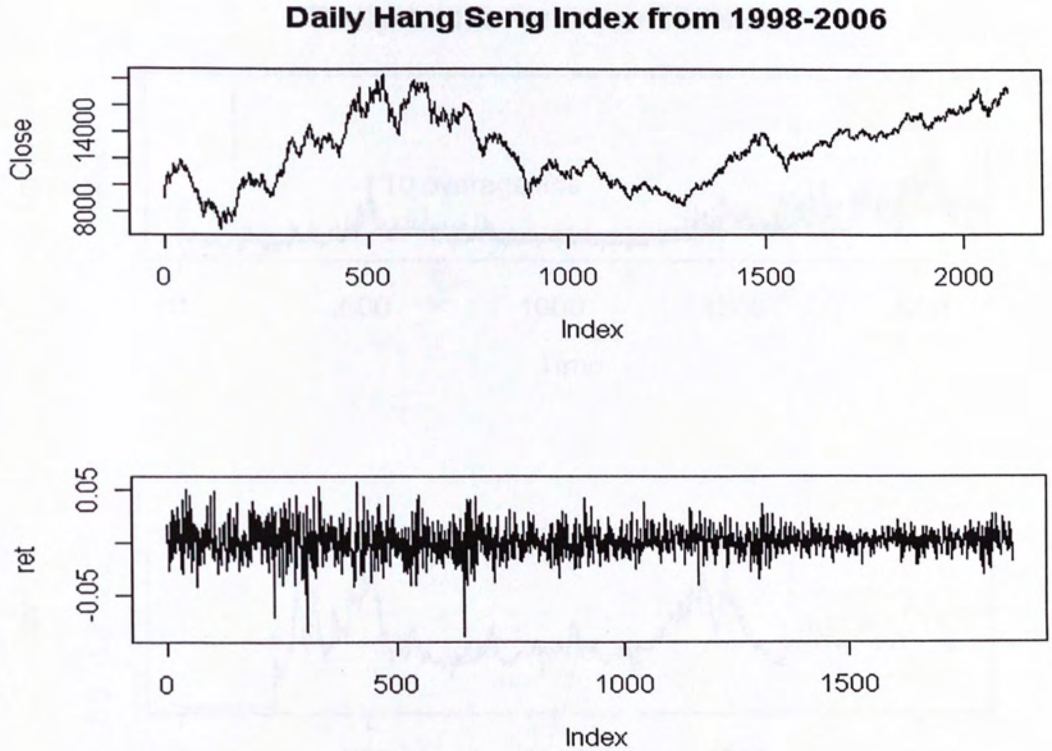

Figure 1: The Daily Hang Seng Index and its return series

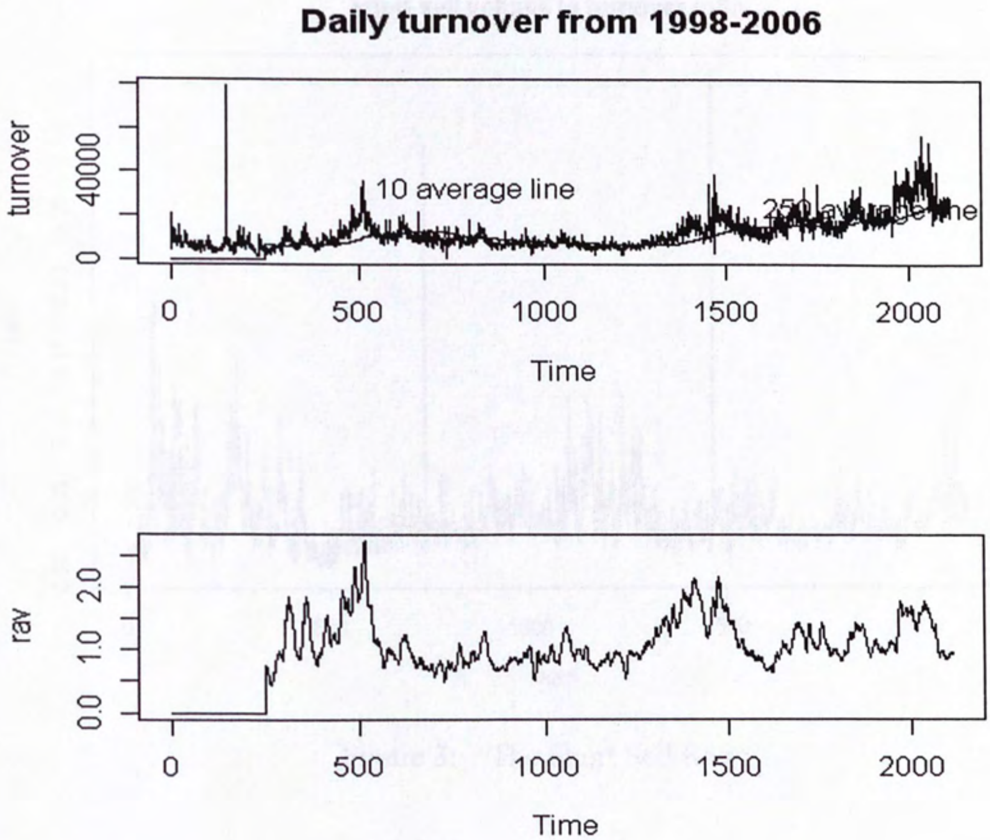

Figure 2: The time series graph for turnover

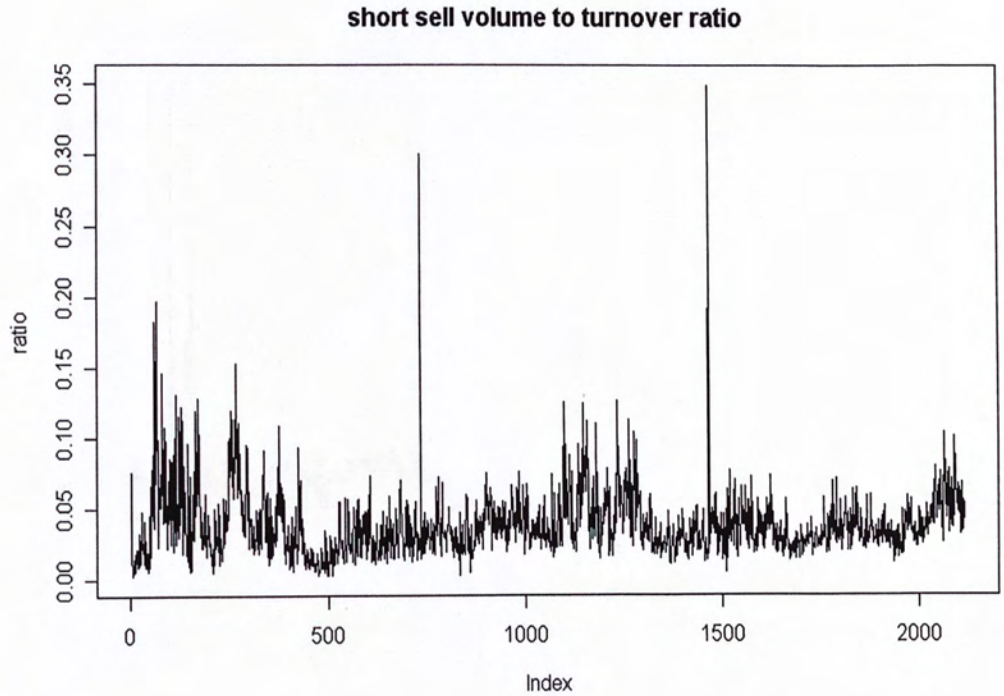

Figure 3: The Short Sell Ratio

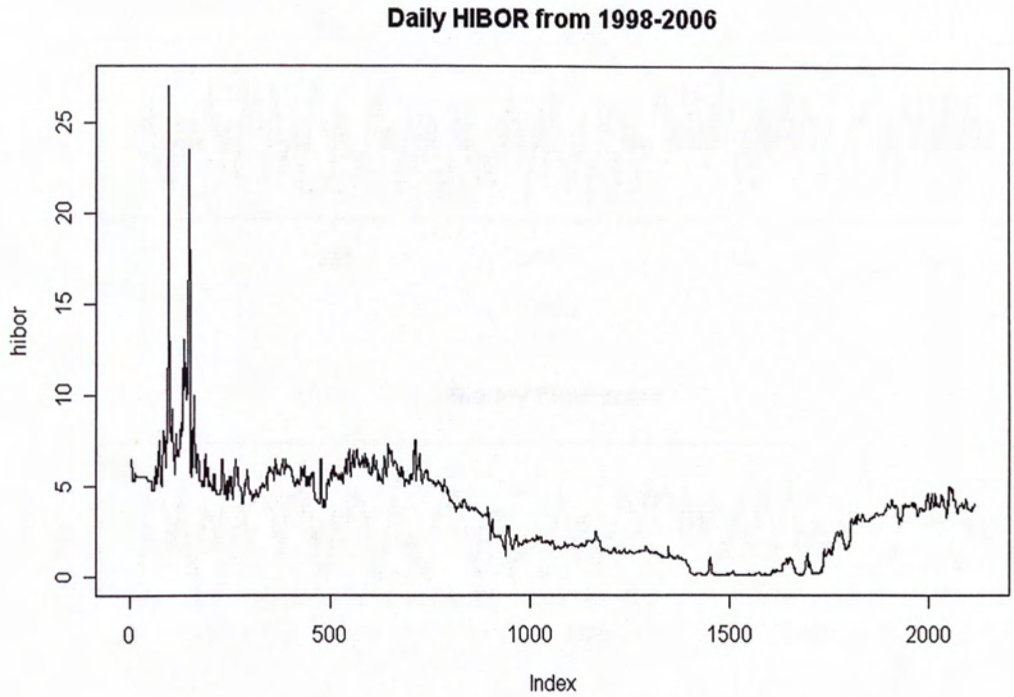

Figure 4: HIBOR

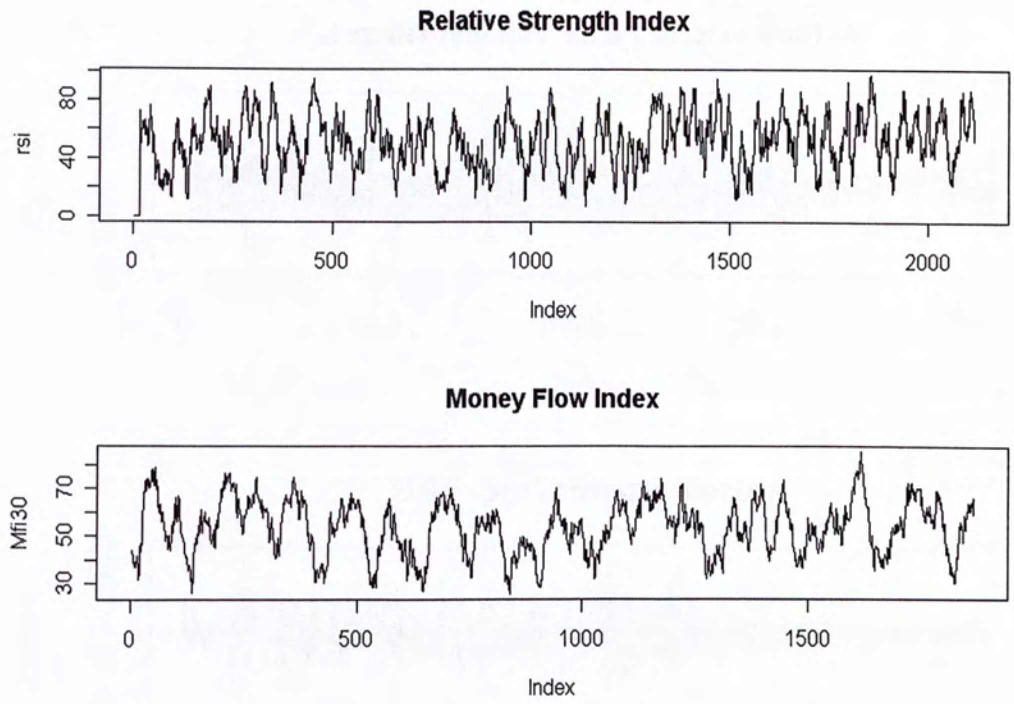

Figure 5 Technical Indicators

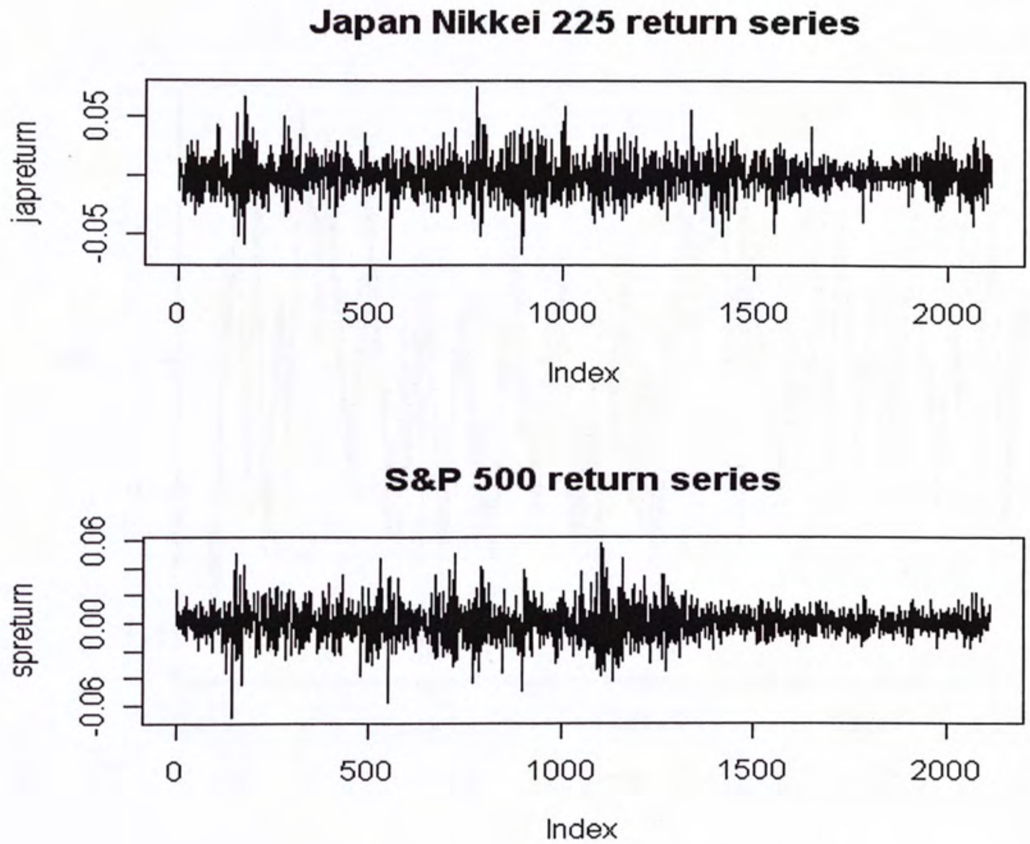

Figure 6 The Japan and USA market index return series

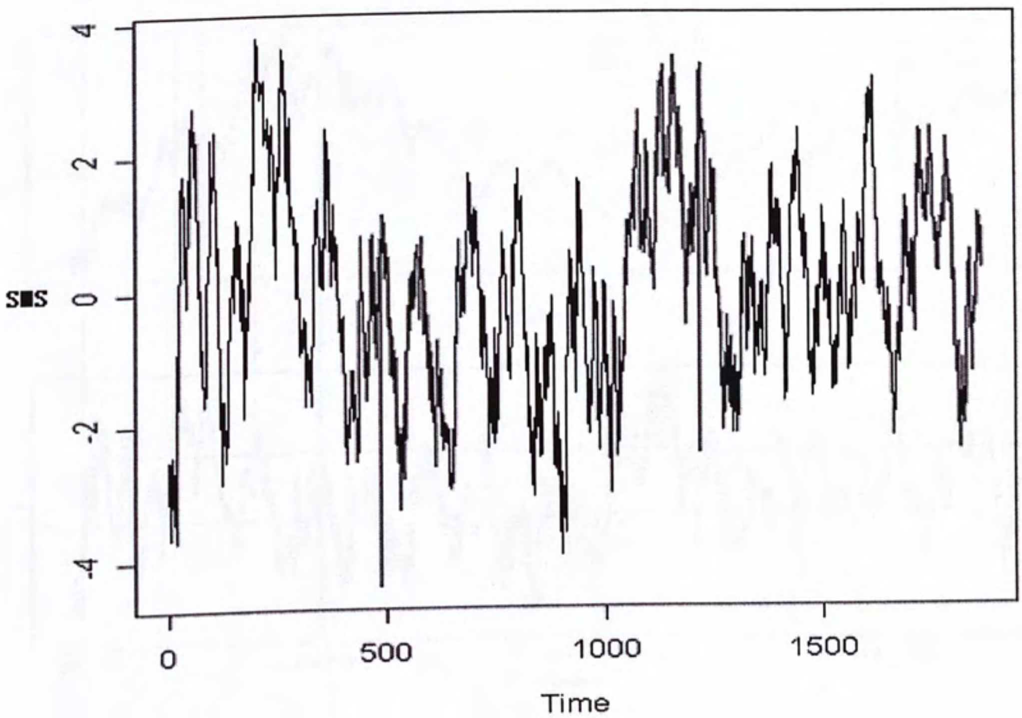

Figure 7   Stock market sentiment Index

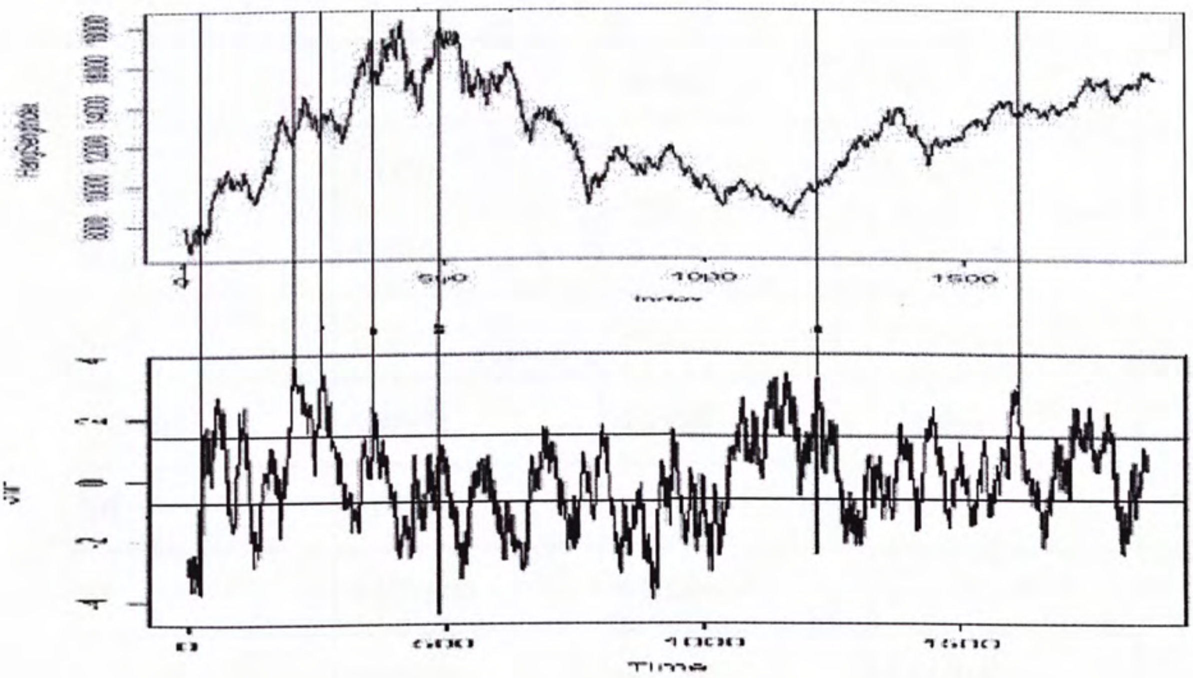

Figure 8   Comparison of SMS and Hang Seng Index

TABLE 1

	Mean	Median	SD
Rav	1.1180	1.0060	0.36780
RSI	52.510	52.390	18.26394
MFI	53.44	54.17	11.1666
HIBOR	0.02974	0.03062	0.02051
SR	0.03704	0.03305	0.02065
SP	9.174e-05	0.000e+00	0.01157247
JAP	0.00016	0.00000	0.014043